文 献 与 数 据 学报
第 5 卷　第 3 期　　2023 年 9 月

数据融合研究的主题与方法趋势*

李　杰[1, 2]　于倩倩[1]　王玉菊[1]
（1. 中国科学院文献情报中心，北京　100090；
2. 中国科学院大学经济与管理学院信息资源管理系　北京　100190）

摘　要:［目的 / 意义］数据融合是实现多源数据价值的重要途径，全面分析全球数据融合研究的整体主题格局，对当前认识和研究数据融合有重要的科技情报价值。［方法 / 过程］采用词频与共词分析法，对 Web of Science 核心数据集中 16 053 篇数据融合研究论文的热点主题和研究方法进行了分析。［结果 / 结论］数据融合研究在整体上呈现了显著的增长趋势，且经过 30 余年的发展已经形成了核心的研究热点和数据融合方法。在研究中，传感器（包括无线传感器）数据融合是该领域的研究热点方向。故障诊断、遥感、安全以及智能电网等是数据融合应用的热点场景。卡尔曼滤波法、神经网络、Dempster-Shafer 证据理论以及机器学习（包括深度学习、支持向量机等）等是数据融合的热点方法，且数据融合研究中已经形成了多方法共现协同网络。

关键词：数据融合　信息融合　知识融合　多源数据集成　共词分析　VOSviewer
分类号：G202
DOI：10.31193/SSAP.J.ISSN.2096-6695.2023.03.03

0 引 言

“兼听则明，偏信则暗”。基于单一数据源，揭示研究对象情报特征和客观规律的缺陷长期存在。鉴于数据融合在整合对象情报中的重要应用价值，20 世纪 90 年代，美国军事领域开始关注并研究数据融合问题。20 世纪 80 年代，由美国国防部联合实验室 JDL（Joint Directors of Laboratories）主导，成立数据融合工作组（Data Fusion Working Group）[1]，开始系统开展数据融合的研究。在研究中，JDL 根据军事上对数据融合的需求，将数据融合定义为：把来自多个传感器和信息源的数据加以关联（association）、相关（correlation）以及集成（combination）的过程，以获得准确的位置和身份估计（position & identify estimation），从而全面细致地对态势、威胁和重要性做出评估[2]。JDL 的数据融合概念和模式，在后来的数据融合研究与应用中发挥了重要作用。

* 本文系中国科学院文献情报能力建设专项项目“研发数据组织与分析挖掘的智能技术”（项目编号：E1290002）的研究成果之一。
［作者简介］李杰（ORCID：0000-0002-4096-2795），男，副研究员，博士，Email：lijie2022@mail.las.ac.cn（通讯作者）；于倩倩（ORCID：0000-0001-8777-1171），女，副研究馆员，硕士，Email：yuqianqian@mail.las.ac.cn；王玉菊（ORCID：0000-0003-2539-2218），女，馆员，硕士，Email：wangyj@mail.las.ac.cn。

虽然，数据融合的讨论与研究已经有 30 余年的历史，但随着时代发展，数据的内涵和外延发生了变化。在新的数字化时代下，重新审视数据融合问题很有必要。

当前，数字化、数据化以及数智化程度进一步深化，集成和融合多源数据进行问题解答，成为当前复杂信息环境下数据驱动问题解决的一个新兴热点。知识管理的 data-information-knowledge-wisdom 模型，向我们展示了从数据到智慧的数据赋能过程，数据、信息以及知识的融合，是实现智慧化的基础。通过文献调查发现，在目前的学术研究中，数据融合、信息融合和知识融合界限并不明确。在实际的应用中，学者往往根据个人或所在研究领域学者的偏好而随机使用。有学者对三者进行了比较和讨论，如祝振媛和李广建教授对数据融合、信息融合、知识融合的关联与比较分析发现，数据融合与信息融合的研究内容日益趋同，研究边界并不十分清晰；知识融合主要侧重对文献与知识的融合，具体表现在知识库的建设、知识地图的构建、本体的构建等方面，这些方面更强调语义和资源之间关系的组织与表达[3]。但从知识融合的研究来看，仅仅是融合的对象发生了变化（即将文献数据等与知识关联密切的数据当作知识来融合），其本质与数据融合和信息融合并没有显著的区别。笔者认为，按照 DIKW 模型，数据融合是最底层最基础的融合模式，是信息融合、知识融合以及情报融合的基础。在理论上，信息融合、知识融合以及情报融合是数据高维度的融合，因此数据融合应该包括信息融合、知识融合以及情报融合。

数据融合的技术与方法，已经渗透到了科学研究的方方面面。在天文学领域，科学家通过多源数据与信息融合技术，探索和绘制宇宙画像（例如：首张黑洞照片的合成）。在自动化领域，通过融合多源传感器的数据（例如：温度、图像、声音、振动等等），以实现控制系统的智能化（例如：自动驾驶）。在科技情报领域，各类科技文献情报源的融合赋能，已经成为实现“整体情报观”和“精准情报观”的必由之路。综合集成思想[4]和融合各个领域专家知识的决策思想和方法，已经成为复杂问题决策过程中的重要方法。长期以来，不同专业领域的学者从领域需求的角度出发，对数据融合理论、技术、方法等都进行了相关研究分析。但由于迫切程度不同，各个领域在数据融合实际应用方面又存在一定的差异。在科技情报领域，数据融合发展相对比较缓慢。科技文献数据融合主要以科技论文为核心，与专利、政策、社交媒体等数据类型关联，以从不同视角来研究科学与技术（S-T）、科学与政策（S-P）以及科学与媒体的互动。在医学科技信息的研究中，通过融合病历的各类诊断信息，能有效地发现潜在的健康隐患，进而为“精准医学”“治未病”提供决策依据。除了以上提及到的数据融合场景外，数据融合在时时刻刻地影响着人类生产、生活以及生存，全面认识数据融合研究意义重大。

国内外已有相关学者对多源数据融合的理论、模型方法和应用做了调查和综述分析。如，韩增奇等[5]对信息融合技术进行了全面综述，回顾国内外的发展状况，主要对信息融合的定义、层次结构、经典方法以及应用做了全面的综述，为认识信息融合的研究态势提供了详实的研究基础。于佳会等[6]通过中国知网的数据，对国内多源多维数据融合研究态势，从理论、方法以及应用三个方面进行了分析。Castanedo[7]和 Alofi 等[8]先后对数据融合的技术进行全面梳理和综述，为全面认识数据融合技术和流程提供了全景素材。此外，相关学者也在不同程度上对数据融合的模型、方法和应用等进行了系统综述和讨论[9-10]。通过对以往的数据融合综述的分析，发现传

感器数据是数据融合的核心领域，在该领域已经发展形成了一套较为完整的数据融合流程、模型和若干技术方法。在以往的研究基础上，为了更加全面认识数据融合的研究热点主题和方法群，本文利用国际性的 Web of Science 数据库，全面采集数据融合的研究文献，采用关键词词频和共词分析（co-words）的方法，对全球数据融合的研究态势以及核心的数据融合方法进行研究与分析，以期为我国在复杂信息环境下，基于数据驱动的决策提供科技情报支撑。

1 数据与方法

1.1 数据来源

本文数据来源于 Web of Science 核心合集，子数据库选择了 SCI/SSCI 期刊论文和 CPCI 会议论文数据库。在数据检索时，按照前期的调研结果，在数据融合、信息融合和知识融合缺乏清晰界定的背景下，本文使用“数据融合”来统称三种数据融合提法。在数据检索中分别使用 fusion、integration，aggregation 来作为“融合”的英文表达，进行数据检索。检索时间范围设定为 1900~2021，以获取截止到 2021 年的所有数据融合科技文献数据。最终的数据检索式如下：

（TI=（“data fusion*” OR “data integrat*” OR “data aggregat*” OR “information fusion*” OR “information integrat*” OR “information aggregat*” OR “knowledge fusion*” OR “knowledge integrat*” OR “knowledge aggregat*”）AND FPY=1900–2021）AND LA=（English）

通过数据检索，共得到 1966~2021 年间发表的 16 053 篇数据融合研究论文，产出趋势如图 1。从论文产出的阶段特征来看，可以将数据融合的发展划分为四个时期：萌芽期、缓慢发展期、快速发展期（I）和快速发展期（II）。1966~1989 年数据融合的研究关注度很低，这与当时数据融合的需求和数字化的发展程度比较低有关。该阶段整体上数据处理规模比较小，单一数据源的分析尚且处于小数据时代，对多源数据融合的需求很少。1990~1998 年，数据融合的论文产出要比上一个阶段活跃，但整体产出量仍然处于比较低的水平，甚至在 1999 年出现了小幅度的下降。

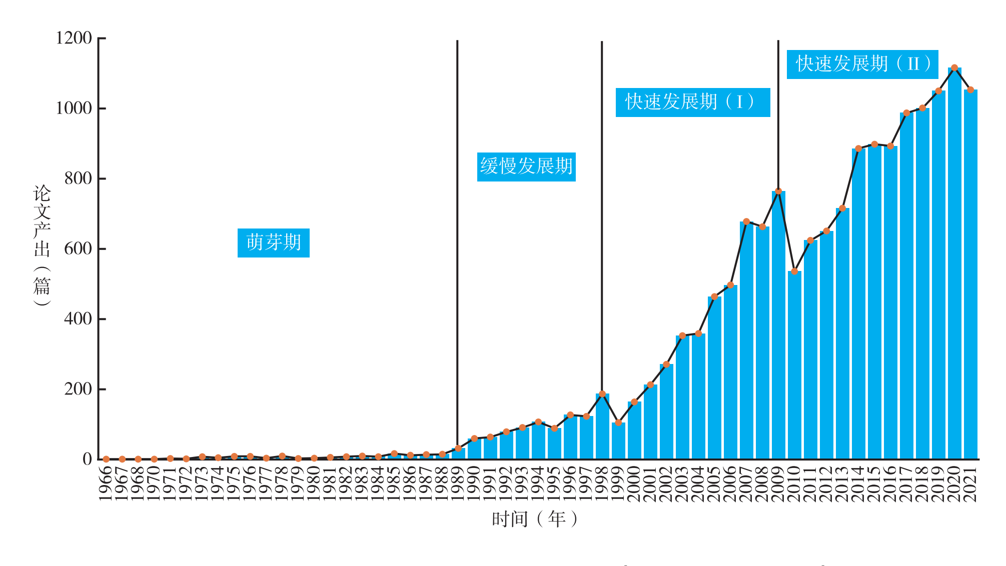


图 1 全球数据融合科研论文产出趋势（1966~2021 年）

1999~2009 年，数据融合的论文产出量进入快速发展期（I），呈现显著增长趋势，表明该阶段对数据融合的需求和关注增加。论文产出在 2010 年经过小幅度下降后，又开始了快速的增长，进入快速发展期（II）。2010~2021 年，大数据、人工智能、物联网等数据采集、数据处理和分析技术发展迅速，数据融合的技术和实际应用需求有了空前的增加，数据融合的科研产出又迎来了新的增长期。

1.2 研究方法

为了探索数据融合的研究主题特征，本研究从词频和共词两个维度对数据融合的主题进行分析。词频统计分析是指以所采集的数据融合论文数据为样本，从数据集合的关键词字段（DE）中，提取并统计每一个关键词出现的次数，以出现次数来衡量关键词的热度。在科技文献数据分析中，通常使用高频的关键词来表征文献数据集的研究热点。关键词词频分析仅仅从单维度呈现文献数据集的研究内容，缺乏对关键词与关键词之间语义关系的揭示。因此，在词频分析的基础上，进一步通过共词分析的方法，对数据融合的研究主题网络进行分析。

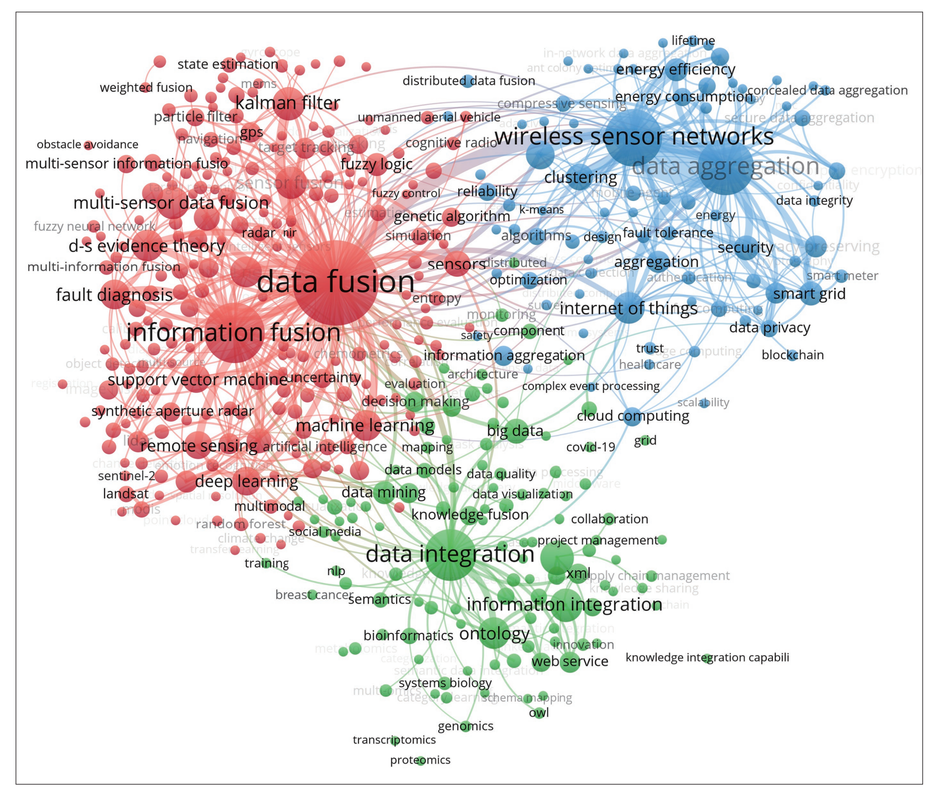
共词分析的提出最早可以追溯到 20 世纪 80 年代，以法国科学研究中心的 Callon 等人[11]出版的《科学技术动态图谱》为代表。随着共词使用的不断广泛，其基本的分析流程和模式已经形成。首先，需要通过文本挖掘或提取技术，识别有意义的术语；在用户定义的“关联或共现”规则的约束下，构建术语与术语之间的关系矩阵；然后，对所生成的共词矩阵进行统计学或数学处理，以解释研究领域的主题特征。早期的共词矩阵常常使用 SPSS 来进行处理，以完成词对在二维空间的映射和词矩阵的层次聚类。近年来，数据库技术和知识图谱工具发展迅速，使得共词分析更加便捷。例如，近年来兴起的 CiteSpace、VOSviewer 以及 SCIMAT 软件等，特别是 VOSviewer 整合了主题词的映射和聚类技术，很大程度上提高了共词分析的效率和效果。因此，本研究选取数据融合论文的关键词作为分析的对象，使用 VOSviewer 科学知识图谱工具对关键词的词频和共词网络进行分析[12-13]。

2 数据融合的热点主题与方法分析

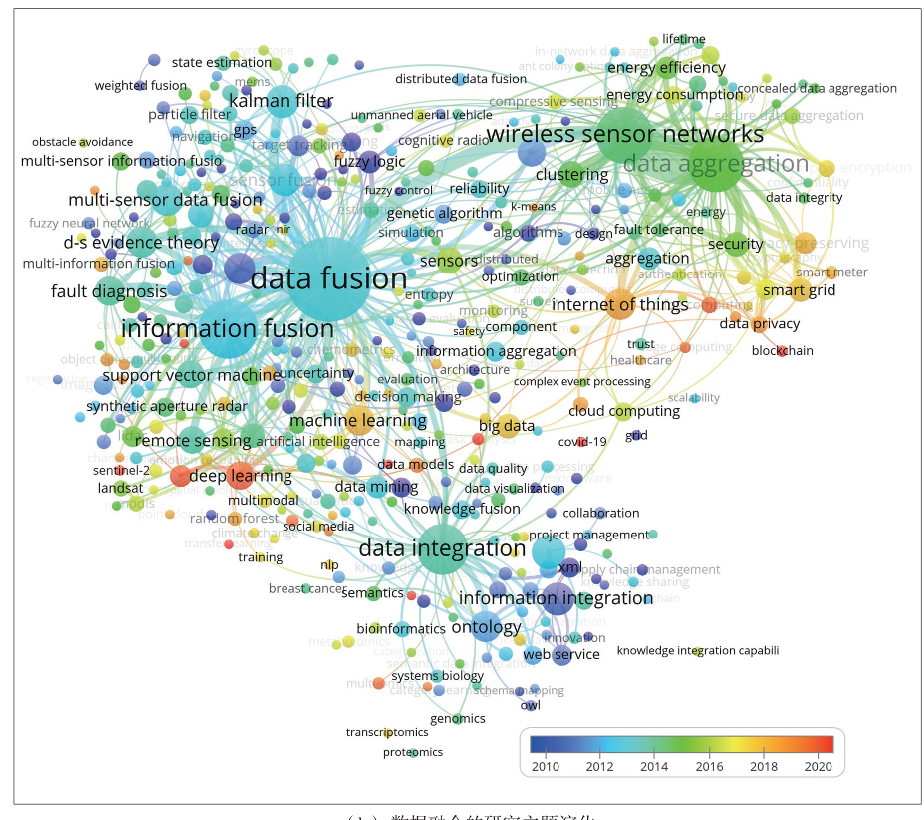
2.1 数据融合热点主题的整体格局

关键词直接表征了论文的研究内容，是通过文献数据进行科学研究热点分析的重要元素。本研究从 16 053 篇论文中提取词频不小于 10 次的 418 个关键词，构建了数据融合主题网络，如图 2（显示了 TOP 500 的连线）。本研究对关键词进行了清洗和消歧处理，即删除了无意义的关键词，合并了异形同义等关键词；为了了解数据融合不同形式所使用的场景偏好，在分析过程中未对检索式中表征数据融合的同义关键词进行合并。

数据融合关键词的词频分析结果显示：除了检索词以外，词频不小于 100 次的热点关键词为 wireless sensor networks（无线传感器网络，872 次）、kalman filter（卡尔曼滤波法，228 次）、multi-sensor data fusion（多传感器数据融合，213 次）、sensor fusion（传感融合，209 次）、neural network（神经网络，205 次）、D-S evidence theory（D-S 证据理论，201 次）、internet of things（物联网，195 次）、ontology（本体，195 次）、machine learning（机器学习，180 次）、fault diagnosis（故障诊断，174 次）、



（a）数据融合的研究主题聚类



（b）数据融合的研究主题演化

图 2 数据融合研究主题的共现网络

sensor networks（传感器网络，164 次）、deep learning（深度学习，152 次）、support vector machine（支持向量机，148 次）、remote sensing（遥感，144 次）、clustering（聚类，135 次）、sensors（传感器，135 次）、multi-sensor data fusion（多传感器数据融合，127 次）、multi-sensor（多传感器，124 次）、classification（分类，116 次）、big data（大数据，114 次）、security（安全，112 次）以及 smart grid（智能电网，102 次）。从这些高频关键词的结果来看，传感器数据融合是当前研究的热点领域（涉及的高频词有 wireless sensor networks，multi-sensor data fusion，sensor fusion，sensor networks，sensors 以及 multi-sensor）。这是自动化、人工智能、物联网等新兴技术发展的必然。其中，无线传感器网络（以下简称“WSN”）集成了传感器、微机电系统和网络技术，是一种全新的信息获取和处理技术，在军事领域、环境科学、医疗检测以及空间探测等领域都有重要的应用价值[14]。特别是 WSN 还存在一些显著的特点，例如：WSN 数据存在很大的冗余，使得 WSN 很少出现运行的问题，可以便捷和快速地进行信息传输和分享，提升了系统的应急能力和情报的时效性。数据融合的相关场景或领域热点关键词主要为故障诊断、遥感、大数据、安全以及智能电网等。高频关键词在时间维度上的分布如图 2（b）。图中，关键词的节点颜色越接近暖色，则关键词在近期就越活跃。分析结果显示，数据融合无论是在解决问题的场景，还是所采用的技术和方法都是与时俱进的。目前，数据融合关键词群中最为活跃的十大关键词分别为 covid-19（新冠）、task analysis（任务分析）、computational modeling（计算模型）、knowledge graph（知识图谱）、transfer learning（迁移学习）、cameras（摄像机）、blockchain（区块链）、convolutional neural network（卷积神经网络）、sentinel-2（哨兵 2 号高分辨率多光谱成像卫星）以及 deep learning（深度学习）。

从共词网络的角度对主题进行聚类分析，结果如表 1。分析表明，虽然中文将 fusion、integration 以及 aggregation 都译为“融合”，但聚类的结果表明它们之间在使用的场景上存在显著的偏好。当前的关键词聚类，主要以检索词为各类中的代表词，将研究划分为 #1 data fusion，#2 data integration 和 #3 data aggregation。聚类 #1 和聚类 #3 所呈现的研究都在关注传感器数据的融合，即传感器数据融合的研究中多使用 data fusion 和 data aggregation 来表征数据融合。虽然 #1 和 #3 都是传感器数据融合，但也存在一定的差异。其中，#3 data aggregation 的关键词群主要集中在 WSN 的数据融合研究方向，#1 data fusion 则集中在多源传感器数据的融合。聚类 #2 data integration 的关键词词群则与 #1 和 #3 显著不同，该类表征为网络或文献数据的融合。

表 1　数据融合关键词聚类中的高频词

编号	关键词词群	关键词词频（词频 = 论文数）
1	#1 data fusion（传感器数据与信息融合）	主题词：data fusion（数据融合，2384）、information fusion（信息融合，988）、multi-sensor data fusion（多源传感器数据融合，213）、sensor fusion（传感器融合，209）、fault diagnosis（故障诊断，174）、remote sensing（遥感，144）、sensors（传感器，135）、multi-sensor（多传感器，124）、feature extraction（特征提取，77）、lidar（激光雷达，75）。 方法词：kalman filter（卡尔曼滤波，228）、neural network（神经网络，205）、D-S evidence theory（D-S 证据理论，201）、machine learning（机器学习，180）、deep learning（深度学习，152）、support vector machine（支持向量机，148）、classification（分类，116）、fuzzy logic（模糊逻辑，90）、convolutional neural network（卷积神经网络，80）。

续表

编号	关键词词群	关键词词频（词频 = 论文数）
2	#2 data integration（网络、知识数据融合）	主题词：data integration（数据融合，688）、knowledge integration（知识融合，251）、information integration（信息融合，231）、ontology（本体，195）、web service（网络服务，73）、xml（可扩展标记语言，65）、decision making（决策，63）、semantic web（语义网络，61）、knowledge fusion（知识融合，56）、knowledge management（知识管理，51）、interoperability（互操作性，50）、multi-agent systems（多主体系统，49）、database（数据库，48）、information retrieval（信息检索，40）、service-oriented architecture（面向服务的体系结构，38）、bioinformatics（生物信息学，37）、component（成分，36）。 方法词：big data（大数据，114）、data mining（数据挖掘，90）、GIS（地理信息系统，64）。
3	#3 data aggregation（无线传感器网络数据融合）	主题词：wireless sensor networks（无线传感器网络，872）、data aggregation（数据融合，845）、internet of things（物联网，195）、sensor networks（传感器网络，164）、security（安全，112）、smart grid（智能电网，102）、aggregation（融合，91）、energy efficiency（能效，88）、privacy-preserving（隐私保护，84）、privacy（隐私，77）、information aggregation（信息融合，68）、reliability（可靠性，52）、energy consumption（能耗，51）、network lifetime（网络寿命，49）、data privacy（数据隐私，48）、homomorphic encryption（同态加密，48）。 方法词：clustering（聚类，135）、cloud computing（云计算，57）、algorithms（算法，48）、optimization（优化，46）。

2.2　数据融合的热点方法与趋势

数据融合方法的剖析和总结，对实际的数据融合研究有重要指导意义。在以往研究中，Jitendra[2]曾系统性地将传感器数据融合的数学方法总结为：概率数据融合方法；模糊逻辑与可能性理论的数据融合；滤波、目标追踪和运动学数据融合；无中心数据融合系统；成分分析与数据融合以及图像代数数据融合等。在国内，数据融合与信息融合的相关综述性或科学计量类文献，也在不同层面上总结了数据融合的分析方法[5,15]。本部分在关键词整体网络的基础上，进一步提取了 TOP 30 的数据融合方法，如图 3。数据融合研究应用的十大方法为卡尔曼滤波法（228 次）、神经网络（205 次）、D-S 证据理论（201 次）、机器学习（180 次）、深度学习（152 次）、支持向量机（148 次）、聚类（135 次，上位方法）、分类（116 次，上位方法）、大数据（114 次，上位方法）、数据挖掘（90 次，上位方法）以及模糊逻辑（90 次）。从方法应用论文的平均时间来看，模糊逻辑、小波分析、神经网络以及遗传算法等是最早一批数据融合分析方法。从早期方法应用来看，卡尔曼滤波法、证据理论以及支持向量机在数据融合中具有高的使用频次，是数据融合中的热点方法。近期在大数据和人工智能背景下，卷积神经网络、深度学习、雾计算以及机器学习等是新兴的数据融合方法。通过方法的共现网络分析，提取了方法的关系，涉及了方法的隶属关系、协同关系等方法关联模式。图 4 呈现了数据融合的方法协同网络，两个方法之间有连线，则表明两种方法存在同时使用的情况，连线越宽则两个方法在一起使用的频次越高。主要的方法关系对为：深度学习—卷积神经网络、神经网络—模糊逻辑、神经网络—DS 证据理论、深度学习—机器学习、深度学习—特征提取、神经网络—遗传算法、支持向量机—分类、机器学习—神经网络、支持向量机—特征提取，等。

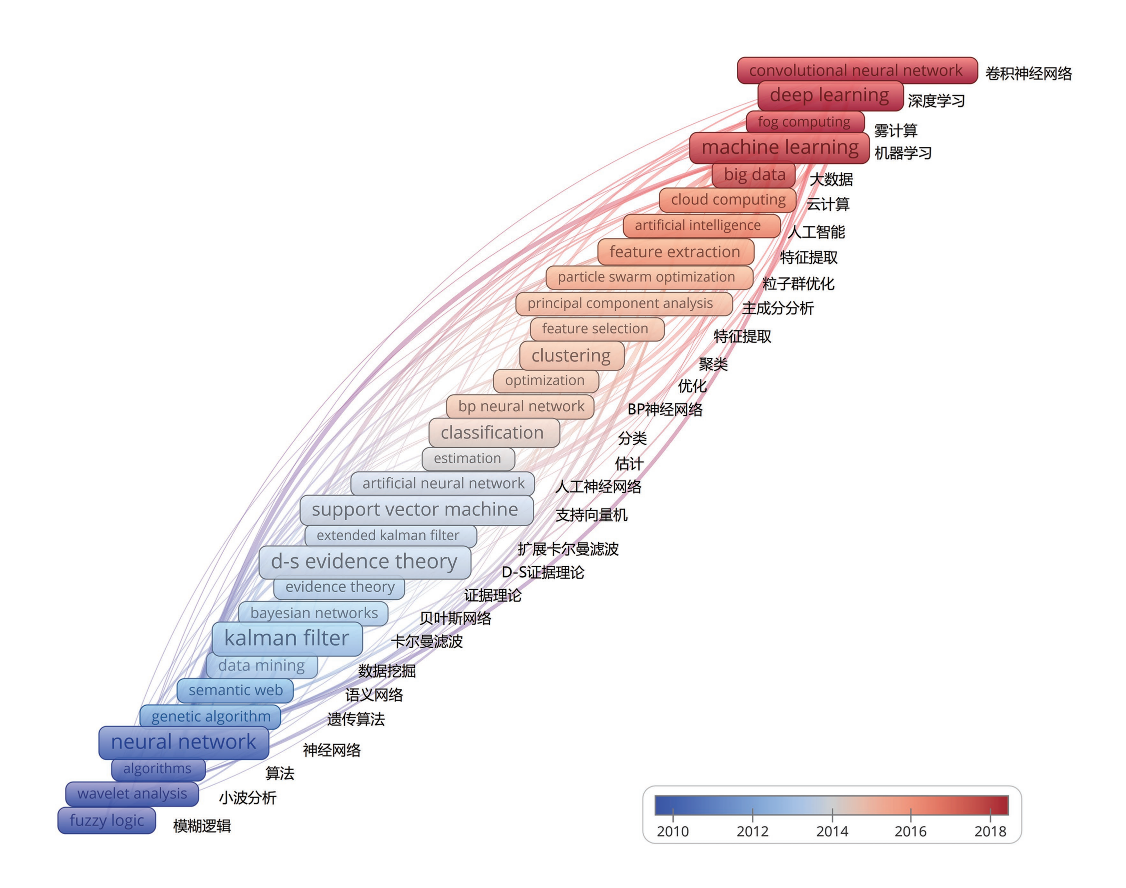


图 3　数据融合方法的时间趋势

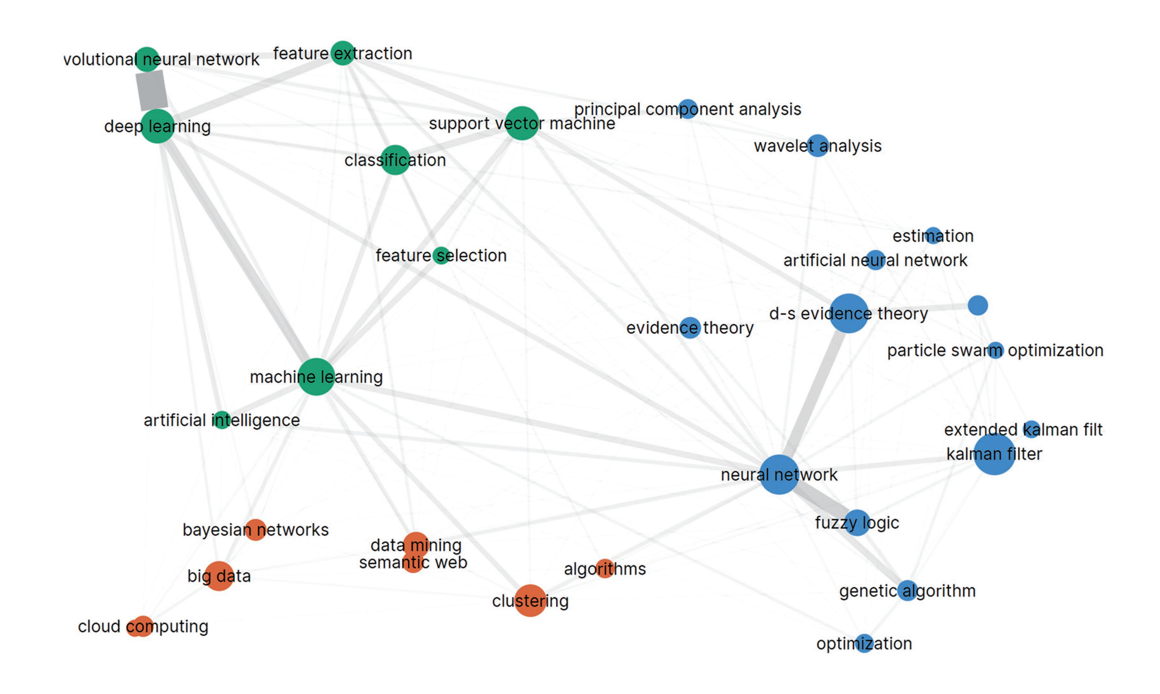


图 4　数据融合方法的协同网络

根据以上分析，选取几种含义较为明确、应用较为广泛的数据融合方法进行介绍：

（1）卡尔曼滤波法（Kalman Filter）

1960 年 Kalman[16] 发表了用递归方法解决离散数据线性滤波问题的论文，后人称其为卡尔曼滤波。该方法的基本思想是：以最小均方误差为最佳估计准则，采用信号与噪声的状态空间模型，利用前一时刻的估计值和当前时刻的观测值来更新对状态变量的估计，求出当前时刻的估计值。算法根据建立的系统方程和观测方程对需要处理的信号做出满足最小均方误差的估计[17]。卡尔曼滤波的特点使其非常适合解决复杂多传感器估计和数据融合问题[2]，且在数据融合的应用中进一步形成了扩展卡尔曼滤波法[18] 和无迹卡尔曼滤波[19] 数据融合方法。当前，卡尔曼滤波法已经广泛地应用在数据融合的研究与实践中。例如，在方法的研究中，Sun 等[20] 提出了一种新的由线性最小方差意义下的矩阵加权多传感器最优信息融合准则，给出了一种具有两层融合结构的通用多传感器最优信息融合分散卡尔曼滤波。Smyth 等[21] 则提出一种多速率卡尔曼滤波方法来解决动态系统测量位移和加速度响应的数据融合问题。在工程数据融合的实例中，涉及的领域包含了航空航天、自动驾驶以及无人系统等方面。

（2）D-S 证据理论（D-S Evidence Theory）

1967 年，哈佛大学数学家 Dempster[22] 在利用上、下概率来解决多值映射问题的研究中，形成了该理论的雏形。在其理论的基础上，1976 年，Shafer[23] 进一步完善和发展了该理论，引入了信任函数（belief functions）的概念，建立了基于“证据”和“组合”来处理不确定性问题的数学方法，并出版了代表性著作《证据的数据理论》（*A Mathematical Theory of Evidence*）。该理论是一种处理不确定性问题的理论，在信息融合、专家系统、模式识别等领域得到了广泛的应用。D-S 证据理论的本质过程实际上是通过数学的方法对证据的融合，是当前数据融合研究中的重要方法和方向之一。如，在实际的应用中，该方法已经被应用到了机器人数据融合[24]、诊断信息融合[25]、可靠性数据融合[26] 以及火灾探测[27] 等方面的研究中。此外，在使用过程中，对该方法的修正[28-30] 和多方法融合[31-32] 也做了相关的研究。

（3）神经网络（Neural Network）、机器学习（Machine Learning）、深度学习（Deep Learning）与支持向量机（Support Vector Machine）

1943 年，心理学家 Mcculloch 和数学家 Pitts[33] 提出了神经网络的相关概念和模型，开创了神经网络研究的先河。神经网络分析的基本思想是模拟人脑进行处理信息的方式，将信息的处理过程分为输入层，中间层（隐藏层）和输出层。神经网络自诞生以来，也取得了很大发展。目前，神经网络的种类包含了 BP 神经网络、REF 神经网络以及近年来发展起来的卷积神经网络等类别。神经网络方法用于数据融合研究与分析具有较长的历史。在本研究的数据集中，神经网络用于数据融合可以最早追溯到 1989 年，Whittington 等[34] 使用神经网络方法，对战术和传感器的数据融合问题进行分析，并以现代海军环境为案例进行了分析和研究。基于神经网络的分析方法，已经在数据融合中得到了广泛的应用。如，Chen 等[35] 使用卷积神经网络和朴素贝叶斯的数据融合方法，进行了基于深度学习的裂纹检测研究。Kolanowski 等[36] 使用 Elman 神经网络对多传感器数据融合进行分析研究。此外，基于神经网络的数据融合方法在目标跟踪[37]、故障诊断[38]、遥感[39] 等数据分析中得到了应用。

近期，机器学习和深度学习方法较多地应用在了数据融合研究中，是数据融合方法与技术的新增长点。机器学习是指计算机通过对先验数据集的决策方法进行学习，从而具备了对同类数据进行预测的过程，是人工智能技术的一个重要的研究领域。机器学习按照学习方法可以分为监督学习、非监督学习和半监督学习。深度学习是机器学习发展的一个分支，其目标是使机器和人一样具有分析能力。机器学习能够实现一种自动化的数据聚类、分类或预测机制，因此对于数据融合而言，机器学习能很好地对多源数据进行特征分析和提取，并在学习规则的驱动下进行数据融合。目前，机器学习类的数据融合方法［40-42］，已经被大量应用在各个领域中。如，这些方法已经被应用于腐蚀测试（corrosion testing）［43］、智慧城市安全［44］以及城市大数据融合［45］等研究中。在机器学习的方法家族中，支持向量机是 Cortes 和 Vapnik［46］于 1995 年首次提出的机器学习方法，是机器学习中一类按监督学习方式对数据进行二元分类的广义线性分类器。该方法具有结构简单、泛化能力强等优点。为解决小样本、非线性、高维数据融合提供了一种有效的方法［47］。如，当前支持向量机方法已经在交通信息融合［48］、目标追踪［49］以及遥感数据融合［50］等方面的数据融合中得到了应用。

（4）模糊逻辑（Fuzzy Logic）

1965 年，Zadeh 发表了《模糊集合》论文［51］，标志着模糊数学的诞生。在数据融合中，其表达有多种形式，包括了 fuzzy control（模糊控制）、fuzzy inference（模糊推理）、fuzzy sets（模糊集合）以及 fuzzy theory（模糊理论）等。模糊逻辑对现有数据融合方法具有强大的推动作用，很大程度上提高了数据融合的可实现性。如，利用模糊逻辑可以生成数据融合规则权重，进而服务于数据融合工作的开展。当前，模糊逻辑已经成为数据融合分析的核心方法之一，被大量应用于具体场景下的数据融合任务中。例如，考虑到多传感器观测数据的不确定和不完整性，早在 1994 年，Abdulghafou 等［52］就利用模糊逻辑开发了基于模糊度测度的融合公式，将模糊理论应用在了数据融合的研究中。此外，模糊逻辑也在目标识别［53］、车联网［54］以及移动机器人导航［55］等场景下的数据融合中得到了应用。

（5）小波分析（Wavelet Analysis）

小波分析又称小波变换，是由法国从事石油信号处理的工程师 J. Morlet 于 1974 年首先提出信号分析方法。小波变换是傅里叶变换的一种扩展方法，提供了一种时域 – 频域的表达形式。作为数学的一个研究分支，该方法已经大量应用于信号分析、图像处理和数值分析等领域［56］。特别是在图像数据需求不断增长的背景下，小波分析在图像处理和图像融合中发挥了重要的价值。如，有研究采用小波变化系数，对模糊的飞机图像进行图像融合分析与处理，将模糊的飞机图片转换为清晰的飞机照片［2］。此外，小波变换的方法也被应用于遥感数据融合［57］、多传感器数据融合［58］以及物联网数据融合［59］等方面。

（6）贝叶斯网络（Bayesian Networks）

贝叶斯网络又称信念网络（belief network）或称有向无环图模型（directed acyclic graphical model），是一种概率图模型。作为基于概率数据融合的方法，早期的研究与应用主要聚焦在可靠性的合并［2］。通过先验信任度、条件 PDF（概率密度函数）和后验概率来表示信任 / 信念的程度，然后基于计算后的后验概率，依据贝叶斯规则进行融合，从而推测得到总可靠性［2］。在基

于贝叶斯理论的数据融合实践中，早在 1988 年，Z. Chair 就发表论文，讨论了分布式贝叶斯与分布式数据融合的相关问题。随后，该方法在数据融合中得到了广泛的应用。如：Vechet 等[60]采用该方法对传感器数据进行了融合；Guerriero 等[61]研究了管道泄漏探测中贝叶斯数据融合方法的应用；Trujillo-Barreto 等[62]将该方法应用到了脑电图 / 脑磁图（EEG/MEG）和功能性核磁共振（fMRI）的数据融合中。

（7）语义网（Semantic Web）

语义网又称 Web 3.0，是由万维网（WWW）的创始人 Tim Berners-Lee 于 1998 年提出的概念。语义网思路的核心是通过给万维网上的文档添加能够被计算机所理解的语义，从而使整个互联网成为一个通用的信息交换介质。在语义网中，RDF（资源描述框架）、OWL（网络本体语言）等是其核心方法技术。语义网从某种意义上来讲，就是通过统一的标准来使存放在互联网上的数据能相互关联和集成使用，是互联网数据融合的核心理论。当前语义网被广泛应用于知识的融合中，例如企业知识融合[63]、钻井风险管理的知识融合[64]以及中医与生物医学的知识融合[65]等领域。从其技术特征来看，可以说语义网是以网络为载体的科技文献数据集成与融合的核心技术，是多源文献数据会聚集成融合赋能的关键。

3 总结与讨论

数据融合是当前大数据时代实现多源数据价值的重要途径，也是当前复杂信息环境下科技文献数据研究的重点。数据融合以整体论的思维方式，通过集成多源数据来实现对特定任务与场景的系统揭示。本文对全球数据融合研究的整体主题和方法分析，或许对认识当前数据融合的研究有一定参考价值。

（1）在学术研究中，数据融合有悠久的历史，但受到技术发展阶段和实际需求的影响，数据融合的方向一直未能受到足够的重视和广泛的应用。1999 年和 2010 年，是数据融合研究的两个转折点。1999~2009 年，全球的数据融合研究呈现出了前所未有的研究热度，论文产出呈现出了显著的增长趋势。经历 2010 年短暂的下降后，整体产出又呈现出了显著增长。2010 年以后数据融合的研究在大数据、人工智能、物联网等技术背景下又重新兴起。与此同时，数据驱动型的问题解决中，数据融合已经成为核心的任务之一。

（2）从整个数据融合研究的主题格局来看，按照检索词形成了明显的主题群落。在当前的科学研究中，以传感器的数据融合为主要研究方向，且无线传感器网络数据的融合已经独立形成了一定规模的主题群。从研究的成熟度来看，传感器数据在数据融合中具有悠久的历史，也最为成熟。军事上对传感器数据的融合需求，促进了数据融合事业的发展，也为数据融合积累了理论和技术条件。多源传感器数据融合技术民用以后，又极大地促进了工业和社会生产与社会治理。在新的时代背景下，网络技术发展迅速，并为实现快速的多源数据传输和融合处理提供了可能。当前，数据融合的技术与方法在故障诊断、安全以及智能电网等方面得到广泛应用。在智能化快速发展的背景下，数据融合在数智化的进程中也将发挥越来越重要的作用。

（3）数据融合是对多来源、多种类数据的融合，其任务具有显著的复杂性。在实际的数据

融合工作开展中，已经针对不同的数据融合需求和数据特征，形成了较为成熟的数据融合方法体系。当前，在所有的数据融合方法中，以卡尔曼滤波最具代表。除此之外，D-S 证据理论、聚类 / 分类、支持向量机等方法也在不同的数据融合场景中广泛应用。在近期的数据融合研究中，基于深度学习、机器学习以及卷积神经网络等数据融合方法也已经兴起。在不断的发展过程中，数据融合研究方法也走向了多方法的融合，已经形成数据融合研究的“方法网络”。

（4）虽然本文在数据检索时，未对数据融合、信息融合和知识融合进行区分检索，但从分析的结果来看，它们在应用场景上存在一定的偏好。Fusion 和 Aggregation 在传感器数据融合中使用最为频繁，其中，Fusion 的使用偏向多传感器数据的融合，Aggregation 又更加偏向无线传感器网络数据融合。Integration 的场景则更加偏向传感器数据之外的其他数据融合，例如网络数据、文本知识数据等。这三个词语在语言学上的意义或能从根本上解释其应用的场景，鉴于本文的主要任务，这里不做过多的赘述。

此外，就数据融合的具体场景来看，传感器数据融合研究已经相对成熟，且形成较大的研究规模。因此，在当前科技文献多源数据的融合研究中，可以在一定程度上借鉴其融合的理论、方法与技术，以形成科技文献数据特有的数据融合理论技术体系。在数据作为重要生产要素的背景下，数据要实现价值最大化，数据必须也必将走向融合。

【参考文献】

［1］Hall D L, Llinas J. An introduction to multisensor data fusion［J］. Proceedings of the IEEE, 1997, 85(1): 6-23.

［2］Jitendra R. Raol. 数据融合数学方法：理论与实践［M］. 王刚，贺正洪，王睿等，译. 北京：国防工业出版社，2021.

［3］祝振媛，李广建.“数据—信息—知识”整体视角下的知识融合初探——数据融合、信息融合、知识融合的关联与比较［J］. 情报理论与实践，2017，40（2）：12-8.

［4］于景元，周晓纪. 从综合集成思想到综合集成实践——方法、理论、技术、工程［J］. 管理学报，2005（1）：4-10.

［5］韩增奇，于俊杰，李宁霞，等. 信息融合技术综述［J］. 情报杂志，2010，29（S1）：110-114.

［6］于佳会，刘佳静，郑建明. 多源多维数据融合研究态势：理论、方法与应用［J］. 情报杂志，2022，41（5）：133-138，207.

［7］Castanedo F. A review of data fusion techniques［J］. The Scientific World Journal, 2013, 2013: 704504.

［8］Alofi A, Alghamdi A A, Alahmadi R F, et al. A review of data fusion techniques［J］. International Journal of Computer Applications, 2017, 167: 37-41.

［9］Esteban J, Starr A, Willetts R, et al. A review of data fusion models and architectures: Towards engineering guidelines［J］. Neural Computing & Applications, 2005, 14(4): 273-81.

［10］Lau B P L, Marakkalage S H, Zhou Y, et al. A survey of data fusion in smart city applications［J］. Information Fusion, 2019, 52: 357-374.

［11］Callon M, Rip A, Law J. Mapping the dynamics of science and technology: Sociology of science in the real world［M］. Springer, 1986.

［12］李杰. 科学知识图谱原理及应用［M］. 北京：高等教育出版社，2018.

［13］Van Eck N J, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping［J］. Scientometrics, 2010, 84(2): 523-538.

[14] 任丰原，黄海宁，林闯. 无线传感器网络 [J]. 软件学报，2003（7）：1282–1291.

[15] Khaleghi B, Khamis A, Karray F O, et al. Multisensor data fusion: A review of the state-of-the-art [J]. Information Fusion, 2013, 14(1): 28–44.

[16] Kalman R E. A new approach to linear filtering and prediction problems [J]. Journal of Basic Engineering, 1960, 82(1): 35–45.

[17] 彭丁聪. 卡尔曼滤波的基本原理及应用 [J]. 软件导刊，2009，8（11）：32–34.

[18] Roumeliotis S I, Bekey G A. An Extended Kalman Filter for frequent local and infrequent global sensor data fusion [C] // Proceedings of the Conference on Sensor Fusion and Decentralized Control in Autonomous Robotic Systems, Oct 14–15, 1997, Pittsburgh, Pa. Bellingham: SPIE – Int Soc Optical Engineering, 1997: 11–22.

[19] Cao J J, Fang J C. Fuzzy adaptive unscented Kalman filter for MIMU/MMC/GPS data fusion [C] // Proceedings of the 7th International Conference on Electronic Measurement and Instruments, Aug 16–18, 2005, Beijing, P R China. Hong Kong: International Academic Publishers Ltd, 2005: 3.380–3.386.

[20] Sun S L, Deng Z L. Multi-sensor optimal information fusion Kalman filter [J]. Automatica, 2004, 40(6): 1017–1023.

[21] Smyth A, Wu M. Multi-rate Kalman filtering for the data fusion of displacement and acceleration response measurements in dynamic system monitoring [J]. Mechanical Systems and Signal Processing, 2007, 21(2): 706–723.

[22] Dempster A P. Upper and lower probabilities induced by a multivalued mapping [J]. The Annals of Mathematical Statistics, 1967, 38(2): 325–339, 15.

[23] Shafer G. A mathematical theory of evidence [M]. Princeton University Press, 1976.

[24] Sun Y S, Zheng C L, Ma P. D-S evidence theory and its application in robot information fusion [C] // Proceedings of the International Conference on Information Science, Automation and Material System, May 21–22, 2011, Zhengzhou, P R China. Stafa-Zurich:Trans Tech Publications Ltd , 2011.

[25] Wu S Q, Jiang W L. Research on data fusion fault diagnosis method based on D-S evidence theory [C] // Proceedings of the International Conference on Measuring Technology and Mechatronics Automation, Apr 11–12, 2009, Zhangjiajie, P R China. Los Alamitos: IEEE Computer Soc, 2009.

[26] Liang L Q, Cai Q, Shen Y J, et al. A reliability data fusion method based on improved D-S evidence theory [C] // Proceedings of the 11th International Conference on Reliability, Maintainability and Safety (ICRMS) – Integrating Big Data, Improving Reliability & Serving Personalization, Oct 26–28, 2016, Hangzhou, P R China. New York: IEEE, 2016.

[27] Cai Z S, Chen M S. Application of data fusion technology based on D-S evidence theory in fire detection [C] // Proceedings of the 6th International Conference on Electronics and Information Engineering (ICEIE), Sep 26–27, 2015, Dalian, P R China. Bellingham: Spie-Int Soc Optical Engineering, 2015.

[28] Pan G, Wu L L. Information fusion based on improved D-S evidence theory [C] // Proceedings of the 2nd International Conference on Information Technology and Management Innovation (ICITMI 2013), Jul 23–24, 2013, Zhuhai, P R China. Stafa-Zurich: Trans Tech Publications Ltd , 2013.

[29] Sun R, Huang H Z, Miao Q. Improved information fusion approach based on D-S evidence theory [J]. Journal of Mechanical Science and Technology, 2008, 22(12): 2417–2425.

[30] Zhou Y M, Xu H J, Sun J F, et al. Multisensor data fusion based on modified D-S evidence theory [C] // Proceedings of the International Conference on Computer Modeling, Simulation and Algorithm (CMSA), Apr 22–23, 2018, Beijing, P R China. Paris: Atlantis Press, 2018.

[31] Chen B, Wang J F, Chen S B. Prediction of pulsed GTAW penetration status based on BP neural network and D-S evidence theory information fusion [J]. International Journal of Advanced Manufacturing Technology, 2010, 48(1–4):

83-94.

［32］Ding H, Hou R C, Ding X Q. A data fusion equipment monitoring method based on fuzzy set and improved D-S evidence theory［C］// Proceedings of the 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Jul 29-31, 2017, Guilin, P R China. New York: IEEE, 2017.

［33］Mcculloch W S, Pitts W. A logical calculus of the ideas immanent in nervous activity［J］. The Bulletin of Mathematical Biophysics, 1943, 5(4): 115-133.

［34］Whittington G, Spracklen C T. The application of neural networks to tactical and sensor data fusion problems［C］// Proceedings of the 1989 First IEEE International Conference on Artificial Neural Networks, (Conf Publ No 313), 16-18 Oct. 1989.

［35］Chen F, Jahanshahi M R. NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion［J］. IEEE Transactions on Industrial Electronics, 2018, 65(5): 4392-4400.

［36］Kolanowski K, Świetlicka A, Kapela R, et al. Multisensor data fusion using Elman neural networks［J］. Applied Mathematics and Computation, 2018, 319: 236-244.

［37］Chin L. Application of neural networks in target tracking data fusion［J］. IEEE Transactions on Aerospace and Electronic Systems, 1994, 30(1): 281-287.

［38］Jing L, Wang T, Zhao M, et al. An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fult diagnosis of planetary gearbox［J］. Sensors, 2017, 17(2): 414.

［39］Cheng L, Wang L, Feng R, et al. Remote sensing and social sensing data fusion for fine-resolution population mapping with a multi-model neural network［J］. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5973-5987.

［40］Birgersson M, Hansson G, Franke U. Data integration using machine learning［C］// Proceedings of the 20th IEEE International Enterprise Distributed Object Computing Conference (EDOC), Sep 05-09, 2016, Univ Vienna, Fac Comp Sci, Vienna, Austria. New York: IEEE, 2016.

［41］Dong X L, Rekatsinas T. Data integration and machine learning: A natural synergy［J］. Proceedings of the VLDB Endowment, 2018, 11(12): 2094-2097.

［42］Waltz E. Machine learning (ML) support to information fusion［C］// Proceedings of the Conference on Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII, Apr 15-17, 2019, Baltimore, MD. Bellingham: Spie-Int Soc Optical Engineering, 2019.

［43］Voelker C, Kruschwitz S, Ebell G. A machine learning-based data fusion approach for improved corrosion testing［J］. Surveys in Geophysics, 2020, 41(3): 531-548.

［44］Abid A, Abbas A, Khelifi A, et al. An architectural framework for information integration using machine learning approaches for smart city security profiling［J］. International Journal of Distributed Sensor Networks, 2020, 16(10): 16.

［45］Liu J, Li T R, Xie P, et al. Urban big data fusion based on deep learning: An overview［J］. Information Fusion, 2020, 53: 123-133.

［46］Cortes C, Vapnik V. Support-vector networks［J］. Machine Learning, 1995, 20(3): 273-297.

［47］Luo Y, Wang Y Z, Sun M. A data fusion approach based on parallel support vector machine［C］// Proceedings of the 1st IITA International Joint Conference on Artificial Intelligence, Apr 25-May 26, 2009, Hainan Isl, P R China. Los Alamitos: IEEE Computer Soc, 2009.

［48］Liu H H, Wang X Y, Tan D R, et al. Study on traffic information fusion algorithm based on support vector machines［C］// Proceedings of the 6th International Conference on Intelligent Systems Design and Applications (ISDA 2006), Oct 16-18, 2006, Jinan Univ, Jinan, P R China. Los Alamitos: IEEE Computer Soc, 2006.

［49］Vasuhi S, Vaidehi V, Midhunkrishna P R, et al. Multiple target tracking using support vector machine and data

fusion [C] // Proceedings of the 3rd International Conference on Advanced Computing (ICoAC), Dec 14–16, 2011, Anna Univ, Dept Comp Technol, Chennai, India. New York: IEEE, 2011.

[50] Zhao S H, IEEE. Remote sensing data fusion using support vector machine [C] // Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Sep 20–24, 2004, Anchorage, AK. New York: IEEE, 2004.

[51] Zadeh L A. Fuzzy sets [J] . Information and Control, 1965, 8(3): 338–353.

[52] Abdulghafour M, Fellah A, Abidi M A, et al. Fuzzy logic–based data integration: Theory and applications [C] // Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Oct 02–05, 1994, Las Vegas, Nv. New York: IEEE, 1994.

[53] Chen B H, Lin C F, Thomopoulos S C A. Fuzzy logic information fusion for object recognition [C] // Proceedings of the Conference on Applications of Fuzzy Logic Technology II, Apr 19–21, 1995, Orlando, Fl. Bellingham: Spie – Int Soc Optical Engineering, 1995.

[54] Tal I, Muntean G M. Using fuzzy logic for data aggregation in vehicular networks [C] // Proceedings of the 16th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications, Oct 25–27, 2012, Dublin, Ireland. New York: IEEE, 2012.

[55] Parasuraman S, Shirinzadeh B. Fuzzy logic based sensors data fusion for mobile robot navigation [C] // Proceedings of the International Conference on Materials, Mechatronics and Automation (ICMMA 2011), Jan 15–16, 2011, Melbourne, Australia. Durnten–Zurich: Trans Tech Publications Ltd, 2011.

[56] Hariharan G. Wavelet analysis—An overview [M] . Wavelet Solutions for Reaction - Diffusion Problems in Science and Engineering. Singapore: Springer Singapore. 2019: 15–31.

[57] Ranchin T. Wavelets for modeling and data fusion in remote sensing [C] // Proceedings of the Conference of the NATO–Advanced–Study–Institute on Multisensor Data Fusion, Jun 25–Jul 07, 2000, Pitochry, Scotland. Dordrecht: Springer, 2002.

[58] Xu L J, Zhang J Q, Yan Y, et al. A wavelet–based multi–sensor data fusion algorithm [C] // Proceedings of the 20th IEEE Instrumentation and Measurement Technology Conference, May 20–22, 2003, Vail, Co. New York: IEEE, 2003.

[59] Guidi B, De Salve A, Ricci L. A data aggregation strategy based on Wavelet for the Internet of Things [C] // Proceedings of the 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Sep 21–24, 2017, Timisoara, Romania. New York: IEEE, 2017.

[60] Vechet S, Krejsa J, Ondrousek V, et al. Sensors data fusion via bayesian filter [C] // Proceedings of the 14th International Power Electronics and Motion Control Conference (EPE–PEMC), Sep 06–08, 2010, Ohrid, Macedonia. New York: IEEE, 2010.

[61] Guerriero M, Wheeler F, Koste G, et al. Bayesian data fusion for pipeline leak detection [C] // Proceedings of the 19th International Conference on Information Fusion (FUSION), Jul 05–08, 2016, Heidelberg, Germany. New York: IEEE, 2016.

[62] Trujillo–Barreto N J, Martinez–Montes E, Valdes–Sosa P A, et al. Bayesian model for EEG/MEG and fMRI data fusion [J] . NeuroImage, 2001, 13(6): S270.

[63] Gu W D, Xia G P, You W J. Enterprise knowledge integration by semantic web [C] // Proceedings of the International Conference on Research and Practical Issues of Enterprise Information Systems, Apr 24–26, 2006, Vienna, Austria. New York: Springer, 2006.

[64] Xu Y Z. Research of knowledge integration based on semantic web for drilling risk management [C] // Proceedings of the 3rd International Conference on Information Computing and Applications (ICICA 2012), Sep 14–16, 2012, Chengde, P R China. Berlin: Springer–Verlag Berlin, 2012.

［65］Yu T, Liu J, Yang S, et al. Semantic web for knowledge integration between traditional Chinese medicine and biomedicine［C］// Proceedings of the 7th International Conference on Information Technology in Medicine and Education (ITME), Nov 13-15, 2015, Anhui, P R China. New York: IEEE, 2015.

Trends of Topics and Methods in Data Fusion Research

Li Jie[1,2]　Yu Qianqian[1]　Wang Yuju[1]

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;
2. Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract:［**Purpose/significance**］Data fusion is an important way to realize multi-source data value. Comprehensive analysis of the overall topics of global data fusion research has an important scientific and technological information value for the current data fusion research.［**Method/process**］The hot topics and research methods of 16053 literatures from Web of Science core collections were analyzed by word-frequency and co-word analysis.［**Result/conclusion**］The data fusion research has shown a significant growth trend, and after more than 30 years development, core research hotspots and methods of data fusion have been formed. In the research, the data fusion of sensors（including wireless sensors）is the core research direction in this field. Fault diagnosis, remote sensing, security and smart grid are the hotspots of the data fusion scenario. Kalman Filter, Neural Network, Dempster-Shafer Evidence Theory and Machine Learning（including Deep Learning, Support Vector Machine, etc.）are the main methods in data fusion, and the synergy network of methods have been formed in data fusion.

Keywords: Data fusion; Information fusion; Knowledge fusion; Multi-source data integration; Co-word analysis; VoSviewer

（本文责编：周　霞）